1   COOLEY LLP
    BOBBY GHAJAR (198719)
2   (bghajar@cooley.com)
    COLETTE GHAZARIAN (322235)
3   (cghazarian@cooley.com)
    1333 2nd Street, Suite 400
4   Santa Monica, CA 90401
    Telephone: (310) 883-6400
5
    KATHLEEN HARTNETT (314267)
6   (khartnett@cooley.com)
    MARK WEINSTEIN (193043)
7   (mweinstein@cooley.com)
    JUDD LAUTER (290945)
8   (jlauter@cooley.com)
    ELIZABETH L. STAMESHKIN (260865)
9   (lstameshkin@cooley.com)
    3175 Hanover Street
10 Palo Alto, CA 94304
    Telephone: (650) 843-5000
11
    CLEARY GOTTLIEB STEEN & HAMILTON LLP
12 ANGELA L. DUNNING (212047)
    (adunning@cgsh.com)
13 1841 Page Mill Road, Suite 250
    Palo Alto, CA 94304
14 Telephone: (650) 815-4131

15 PAUL, WEISS, RIFKIND, WHARTON & GARRISON LLP
    KANNON K. SHANMUGAM (*pro hac vice*)
16 (kshanmugam@paulweiss.com)
    2001 K Street, NW
17 Washington, DC 20006
    Telephone: (202) 223-7300

    [Full Listing on Signature Page]
18 *Counsel for Defendant Meta Platforms, Inc.*

19   UNITED STATES DISTRICT COURT

20   NORTHERN DISTRICT OF CALIFORNIA

21   SAN FRANCISCO DIVISION

22 RICHARD KADREY *et al.*,   Case No. 3:23-cv-03417-VC-TSH

23   Individual and Representative Plaintiffs, **DEFENDANT META PLATFORMS, INC.'S:
                                                 REPLY IN SUPPORT OF ITS MOTION FOR
24   v.   PARTIAL SUMMARY JUDGMENT**

25 META PLATFORMS, INC., a Delaware   Date: May 1, 2025
    corporation,   Time: 10:00 a.m.
26   Dept: Courtroom 4 – 17th Floor
                               Defendant.   Judge: Vince Chhabria
27   Trial Date: None
                                                 Date Action Filed: July 7, 2023
28

                                                 **META'S REPLY ISO MOT. FOR PARTIAL
                                                 SUMMARY JUDGMENT
                                                 3:23-CV-03417-VC**

TABLE OF CONTENTS

1

TABLE OF AUTHORITIES

2

Page(s)

3

Cases

**META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC-TSH**

**META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC**

TABLE OF EXHIBITS

Ex. No.	DESCRIPTION
EXHIBITS TO REPLY DECLARATION OF BOBBY GHAJAR ("BG II")[1]

Ex. No.	DESCRIPTION
54	Excerpts from the transcript of the December 10, 2024 deposition of Plaintiff Klam, which reflect his testimony as to whether he has used Meta's Llama models, together with excerpts of certain Plaintiffs' Responses to Meta's Requests for Admissions, namely, Responses to Request No. 25 (No. 27 as to Pl. Farnsworth), concerning Plaintiffs' personal use, if any, of Meta's Llama models
55	Excerpts from the transcript of the deposition of Plaintiffs' rebuttal expert Dr. David Choffnes, taken March 28, 2025
56	Excerpts from the transcript of the deposition of Plaintiffs' expert Dr. Cristina Lopes, taken February 13, 2025
57	Excerpts from the transcript of the second deposition of Meta's expert Barbara Frederiksen-Cross, taken April 4, 2025
58	Rebuttal Expert Report of Barbara Frederiksen-Cross, dated February 10, 2025
59	Second Rebuttal Expert Report of Barbara Frederiksen-Cross (to the "Rebuttal Report of David R. Choffnes, Ph.D, February 26, 2025"), dated April 1, 2025
60	Excerpts from the transcript of the Rule 30(b)(1) deposition of Nikolay Bashlykov, taken December 5, 2024
61	Excerpts from the transcript of the Rule 30(b)(1) deposition of Alex Boesenberg, taken November 18, 2024
62	Excerpts from the transcript of the Rule 30(b)(6) deposition of Sy Choudhury, taken December 5, 2024

---

[1] For the Court's ease of reference, exhibits attached to the BG II declaration submitted in support of this reply start at No. 54; citations to BG Exs. 1–53 are to the exhibits to the BG Declaration (Dkt. 490) submitted in support of Meta's Motion for Partial Summary Judgment and Opposition to Plaintiffs' Motion for Partial Summary Judgment [Dkt. 489] ("Meta Mot."). Meta refers to Plaintiffs' summary judgment motion [Dkt. 472] as "Pls' Mot." and to Plaintiffs' Reply to Motion for Partial Summary Judgment and Opposition to Meta's Motion for Partial Summary Judgment [Dkt. 517] as "Opp." Meta refers to the amicus briefs filed in support of Meta as follows: Brief of Intellectual Property Law Professors [Dkt. 509-2] ("IP Profs.' Br."); and Brief of Electronic Frontier Foundation [Dkt. 516] ("EFF Br."). Meta refers to the amicus briefs filed in support of Plaintiffs as follows: Brief of Copyright Law Professors [Dkt. 525] ("Pls' Profs.' Br."); Brief of International Association of Scientific, Technical and Medical Publishers [Dkt. 527] ("IASTM Br."); Brief of Copyright Alliance [Dkt. 529] ("CA Br."); and Brief of Association of American Publishers [Dkt. 535] ("AAP Br."). Unless otherwise noted, all other abbreviations used herein are as defined in Meta's Motion, all emphasis is added, and internal citations and quotation marks are omitted.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

Ex. No.	Description
63	Excerpts from the transcript of the Rule 30(b)(1) deposition of Melanie Kambadur, taken September 17, 2024
64	Declaration of ███████, Senior Director of Strategy & Innovation in the Office of the CEO at ██████, dated April 4, 2025
Exhibits to Reply Declaration of Prof. Michael Sinkinson, Ph.D. ("Sinkinson II")	
	No exhibits
Exhibits to Reply Declaration of Joelle Pineau ("Pineau")	
	No exhibits
Exhibits to Reply Declaration of Melanie Kambadur ("Kambadur")	
	No exhibits

I.   INTRODUCTION

The aims of copyright are to encourage creativity, expand human knowledge, and enrich our shared culture. Thus, "the Supreme Court has long held that copyright law is intended to benefit the public, not authors." *Thaler v. Perlmutter*, 130 F.4th 1039, 1049 (D.C. Cir. 2025). That is why anyone can read and learn from a book without permission—whether they buy it new (and the author is remunerated), or instead buy it used, borrow it from a friend, or find it on a park bench (in which cases the author is not). That is also why the Copyright Act allows the unauthorized copying of a work to transform it into something new; such uses serve distinct purposes and result in new expression that expands the public discourse. Meta's Llama models epitomize these objectives, providing a tool enabling anyone to create new, non-infringing text, benefiting users and the public without substituting for the works on which they were trained.

Plaintiffs fail to meaningfully engage with the undisputed facts and bedrock legal principles establishing that Meta's use of Plaintiffs' books to develop and train Llama was fair. They have no evidence to rebut that Llama serves a fundamentally different purpose than their books, without causing Plaintiffs any harm cognizable under copyright. After nearly two years of litigation, there still is no evidence that anyone has ever used Llama as a substitute for reading Plaintiffs' books, or that they even could. Llama does not output any of Plaintiffs' books, or allow anyone to read them. Instead, Plaintiffs merely repeat their refrain that Meta copied their work for free and without permission—ignoring the many Supreme Court and circuit cases holding that the fair use doctrine *permits exactly that* where, as here, the result of such copying is a new technology that indisputably does not reproduce their expression. Llama is the textbook definition of transformative fair use.

Plaintiffs urge the Court to disregard the fair use factors altogether, insisting that "fair use cannot apply if the copyrighted work was illegally obtained in the first place." Opp. at 4. Meta addressed this argument in its Motion (at 18–21), but Plaintiffs *still* cite no valid authority for this proposition, and it is wrong, *see Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577–78 (1994); *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 32 (2021). Indeed, Sony made this exact argument in *Connectix*, urging that the defendant engaged in bad faith precluding fair use because it had downloaded a bootleg copy of Sony's software to create a competing game system, yet the Ninth

1   Circuit analyzed all four factors and found use of the software to be fair "as a matter of law." *Sony*

2   *Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 608 (9th Cir. 2000). Likewise here, the fair

3   use factors would weigh decisively in Meta's favor *even if* bad faith were a relevant consideration.

4   Plaintiffs and their amici also attempt to avoid a ruling that Meta's *copying* was fair by

5   pointing to unsubstantiated allegations that Meta *distributed* Plaintiffs' works via torrent. But

6   distribution is a separate claim on which no party moved for summary judgment. Nor do Plaintiffs

7   have facts to establish that Meta distributed any of their works. Plaintiffs' baseless distribution theory

8   cannot create a genuine issue precluding summary judgment for Meta that its *copying* was a fair use.

9   Applying well-established law to the undisputed facts of this case, Meta's copying of datasets

10   containing Plaintiffs' works to develop Llama was a quintessentially transformative, non-infringing

11   fair use under 17 U.S.C. § 107. Nor have Plaintiffs presented a coherent theory, much less sufficient

12   evidence, to support their copyright management information ("CMI") removal claim under 17

13   U.S.C. § 1202—ignoring sworn testimony as to the practical reasons why Meta removed redundant

14   information from datasets (none of which was for the purpose, or had the effect, of "concealing"

15   infringement). The Court should enter judgment for Meta on Plaintiffs' §§ 106(1) and 1202 claims.

16   **II.   PLAINTIFFS' "DISTRIBUTION" CLAIM IS NOT PROPERLY BEFORE THE COURT AND DOES
         NOT CREATE A GENUINE ISSUE AS TO WHETHER META'S COPYING WAS FAIR**

17

18   Plaintiffs devote significant attention in their latest brief to arguing that Meta violated the

19   Copyright Act by distributing their works. Opp. at 2–10. But as Meta's Motion noted (at 2 n.1),

20   and Plaintiffs do not dispute, Plaintiffs did not seek summary judgment that Meta distributed their

21   works in violation of § 106(3); they moved only as to Meta's *copying* under § 106(1).[2] Pls' Mot.

22   at 21 ("*copying* of the Books without Plaintiffs' permission violated Plaintiffs' exclusive rights under

23   § 106(1)"); *id.* at 19 ("[T]he Court should grant summary judgment to Plaintiffs on direct

24   _____

[2] As Plaintiffs acknowledge, distinct uses of a copyrighted work "require separate analyses." Opp.
25   at 11. Both uploading and downloading copyrighted works may be infringing if not fair use. Opp.
     at 6–7. "Uploading a copyrighted [] file to other users … violates the copyright holder's § 106(3)
26   distribution right"; "[d]ownloading a copyrighted [] file … violates the copyright holder's § 106(1)
     reproduction right." *Columbia Pictures Indus., Inc. v. Fung*, 2009 WL 6355911, at *8 (C.D. Cal.
     Dec. 21, 2009), *aff'd in part as modified*, 710 F.3d 1020, 1034 (9th Cir. 2013). "The fair use
27   provision, and the first factor in particular, requires an analysis of the specific 'use' of a copyrighted
     work[,]" because it "may be fair when used for one purpose but not another." *Andy Warhol Found.*
28   *for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 533 (2023). Reproduction and distribution
     are, thus, distinct uses that must be evaluated separately.

1   infringement due to Meta's *copying* the Books *from* pirated databases." They also moved for

2   summary judgment that Meta's *copying* was not fair use, but sought no such finding on distribution.

3   *Id.* at vii (Not. of Mot.) (seeking summary judgment that "Meta's *reproduction* of Plaintiffs'

4   Copyrighted Books without permission, including through peer-to-peer file sharing, is not fair use

5   under 17 U.S.C. § 107"), 22 ("Meta's Initial Reproduction of Pirated Copies of Plaintiffs' Books Is

6   Not Fair Use"). No distribution claim is properly before the Court on summary judgment.

7   Plaintiffs nevertheless assert for the first time on reply that "summary judgment is warranted

8   on the basis of Meta's distribution, as well as Meta's copying, of the Books through P2P networks"

9   because "there is no dispute that Meta uploaded massive amounts of copyrighted data, and … made

10   Plaintiffs' Books available for upload[.]" Opp. at 10. In fact, Meta disputes every part of this

11   assertion, and for good reason. First and foremost, Plaintiffs may not seek summary judgment on

12   distribution by way of their reply brief. *Gabriel v. Cnty. of Sonoma*, 725 F. Supp. 3d 1062, 1078–

13   79 (N.D. Cal. 2024) (refusing to address "arguments raised for the first time in [summary judgment]

14   reply brief"); *Arnett v. Seaside Transp. Servs., LLC*, 2015 WL 12711336, at *1 n.1 (N.D. Cal. June

15   3, 2015) (Chhabria, J.) (same). Equally inappropriate is Plaintiffs' argument that "Meta offers no

16   facts to dispute the overwhelming probability that it also uploaded Plaintiffs' works[.]" Opp. at 7.

17   Meta had no burden to refute a motion Plaintiffs did not make.[3]

18   Even if the Court were inclined to consider Plaintiffs' improper distribution arguments, they

19   are of no help to Plaintiffs because Plaintiffs lack evidence to show that Meta distributed any of

20   their works. According to Plaintiffs, torrenting may involve two forms of "data uploading":

21   "leeching" (simultaneously uploading pieces of a file while a download is still underway) and

22   "seeding" (which occurs only after a download completes). Pls' Mot. at 12; Meta Mot. at 34.

23   Plaintiffs have abandoned any claim based on seeding,[4] leaving only potential leeching at issue.

24

25   [3] Although Plaintiffs did not move for summary judgment on distribution, Meta called out the lack of evidence to support a distribution claim and noted that it may seek leave to affirmatively move

26   for summary judgment on any distribution claim once the newly ordered fact and expert discovery was completed. Meta Mot. at 34–35, 2 n.1.

27   [4] Plaintiffs concede that Meta used a "script that worked to prevent 'seeding,'" i.e., that Meta did *not* upload files it had already downloaded via torrent. Pls' Mot. at 13; Meta Mot. at 34. Further,

28   Dr. Choffnes's opinions on the probability of distribution did not take seeding into account. BG II Ex. 55 at 73:15-18; BG II Ex. 59 (04/1/25 Frederiksen Report) ¶¶ 30–31.

1 Opp. at 8. Yet Plaintiffs also admittedly lack evidence that Meta ever uploaded any of their works,

2 or any identifiable part of those works, during the so-called "leeching" phase. Rather, the most they

3 claim their expert can show is a probability that Meta uploaded "at least *some* [unspecified] portion

4 of *one* of Plaintiffs' Books" "*or* at minimum, made them available to peers[.]" *Id.* at 7, 8 (emphases

5 in original). This does not raise a genuine issue of fact as to distribution of Plaintiffs' works.

6 Specifically, Plaintiffs' expert conceded that he had not identified *any* of the more than four

7 dozen at-issue works that was allegedly distributed. BG II Ex. 55 (Choffnes Dep.) at 106:7–15.)

8 Because § 106(3) requires a showing of distribution "of the copyrighted work," and Plaintiffs can

9 identify no work that Meta is supposed to have distributed, they cannot prove this claim. Plaintiffs'

10 argument also fails because, in place of evidence of actual distribution, it relies on a probability

11 analysis that is irremediably flawed. BG II Ex. 59 (4/1/25 Frederiksen Report) ¶¶ 69–81.

12 Distribution of Plaintiffs' works was, in fact, highly *un*likely given the tiny fraction of the

13 downloaded datasets that comprise Plaintiffs' works, the fact that Meta could have uploaded no

14 more than 30% of the data volume it downloaded via torrent, and many other factors that would

15 have had to have aligned for distribution of Plaintiffs' works to have actually occurred. *Id.* ¶¶ 28–

16 60; BG II Ex. 57 (Frederiksen Dep.) 118:12–121:18. And even if Plaintiffs' own flawed analysis

17 were applied to assess the probability that Plaintiffs' works were distributed, the likelihood that

18 Meta distributed a piece from *each* of their works is vanishingly small; for the Internet Archive

19 dataset, for example, that probability is 0.0000000000000000002848%—effectively zero. BG II

20 Ex. 59 ¶¶ 34–37. In short, the record falls far short of establishing distribution of Plaintiffs' works.

21 Recognizing that they cannot prove actual distribution, Plaintiffs argue that they need only

22 show that Meta made works "available" for distribution. Opp at. 6, 9–10. That argument also fails.

23 Plaintiffs' only authority for their assertion that merely "making available" works, without

24 actually uploading them, is sufficient to establish distribution is *Hotaling v. Church of Jesus Christ*

25 *of Latter-Day Saints*, 118 F.3d 199, 203 (4th Cir. 1997). Opp. at 9. But that is not the law in the

26 Ninth Circuit. In *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1162 (9th Cir. 2007), the

27 Ninth Circuit affirmed that "distribution requires an 'actual dissemination' of a copy," i.e., actually

28 "transmitting the [copyrighted work] electronically to the user's computer." Likewise, in *VHT, Inc.*

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   *v. Zillow Group, Inc.*, 918 F.3d 723, 736 (9th Cir. 2019), the Ninth Circuit rejected the "mak[ing]

2   available" theory as "neither supported by the [Copyright Act] nor embraced by this court."

3   Although *VHT* involved a claim for infringement of the exclusive "display" right, district courts in

4   the Ninth Circuit consistently cite it in rejecting *Hotaling's* "making available" distribution theory.

5   *See, e.g.*, *EVOX Prods., LLC v. Verizon Media Inc.*, 2021 WL 3260609, at *3–4 (C.D. Cal. May 5,

6   2021) (citing *VHT* and holding that a distribution claim "based on a 'making available theory' …

7   fails as a matter of law."). Even the single "in-circuit" case Plaintiffs cite (Opp. at 9) as supposedly

8   supporting application of *Hotaling* held just the opposite: that distribution requires "an actual

9   dissemination" and that "making available" does not suffice under *Perfect 10* and *VHT*. *SA Music,*

10   *LLC v. Amazon.com, Inc.*, 2020 WL 3128534, at *2–3 (W.D. Wash. June 12, 2020).; *see also*

11   *Atlantic Recording Corp. v. Howell*, 554 F. Supp. 2d 976, 983 (D. Ariz. 2008) ("Unless a copy of

12   the work changes hands in one of the designated ways, a 'distribution' under § 106(3) has not taken

13   place. Merely making an unauthorized copy of a copyrighted work available to the public does not

14   violate a copyright holder's exclusive right of distribution.")

15   Plaintiffs also cite *Columbia Pictures*, 710 F.3d at 1034, for the proposition that summary

16   judgment can be granted based on statistical evidence showing a probability of distribution. Opp.

17   at 6, 8. But in that case, statistical evidence showing a probability that most of the files uploaded

18   were copyrighted was "corroborated by evidence of specific instances of downloads and transfers

19   of copyrighted works," including a declaration from a third party who had used defendant's site to

20   download one of the plaintiff's copyrighted works and evidence of 400 other uploads to specific IP

21   addresses. *Columbia Pictures*, 2009 WL 6355911, at *9. Plaintiffs also "linked the United States-

22   based downloads (as identified by Plaintiffs' experts) with copyrighted works owned by each of

23   the individual Plaintiffs." *Id.* Plaintiffs admittedly have no such evidence in this case.

24   Finally, Plaintiffs seek to blame Meta for their inability to prove actual dissemination,

25   alleging that Meta "'destroyed' evidence that might have identified content it distributed[.]" Opp

26   at 7. These baseless allegations of spoliation do not excuse Plaintiffs' failure of proof. The parties

27   expressly agreed in the stipulated ESI Order entered by the Court that "[s]ystem, server, or network

28   logs" should "not be preserved" based on the proportionality factors. Dkt. 100-1, at 7. Plaintiffs

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   also explicitly informed the Court in seeking leave to amend their complaint that they would not

2   seek and did not need any further discovery to support their distribution claim. Dkt. 300 at 6–7.[5]In

3   any event, Meta has, in fact, produced logs for the 2022 download and usage records for the 2024

4   download (BG II Decl. ¶ 15), and the 2023 logs Plaintiffs complain were "overridden" (Opp. at 7)

5   are irrelevant because the torrent download Meta performed in 2023 was of academic publications

6   from SciMag, did not include fiction books (or any of Plaintiffs' books), and thus could not have

7   resulted in distribution of Plaintiffs' works. Meta Mot. at 8; Dkt. 494 ¶¶ 4–5; *see* BG II Ex. 58 ¶¶

8   75–77, 85–87.[6] The portion of LibGen downloaded in 2023 that contains Plaintiffs' works (LibGen

9   Fiction) was downloaded via a "direct download" technique that did not involve use of BitTorrent

10   and, thus, did not pose any risk of uploading data to others. Bashlykov Decl. (Dkt. 494) ¶¶ 4–5; BG

11   II Ex. 58 (2/10/25 Frederiksen Report) ¶¶ 75–76.

12   In sum, although Plaintiffs' Opposition asserts that they "*need* not prove actual distribution

13   of their books" to prevail on summary judgment, what it really serves to show is that they *cannot*

14   prove this claim. Opp. at 6–10. Plaintiffs cannot rely on unsupported and untenable allegations

15   that Meta distributed their works—a claim on which they did not move for summary judgment—

16   to defeat Meta's own Motion for summary judgment that its *copying* of Plaintiffs' works was fair.[7]

17   **III.   META'S COPYING TO DEVELOP AND TRAIN LLAMA WAS FAIR USE**

18   Plaintiffs' Opposition fails to rebut Meta's showing that its copying to develop and train

19   Llama was fair use. Contrary to Plaintiffs' assertions, this determination is appropriate at summary

20   judgment under a straightforward application of copyright law to the undisputed facts.

21

22   [5] As Judge Hixson found, Plaintiffs never sought any logs relating to torrenting before the close of
     discovery. *See* Dkt. 400 at 6–7. Plaintiffs sought logs relating to torrenting for the first time in
23   their March 5, 2025 discovery proposal (Dkt. 467), which was adopted by the Court. In accordance
     with the ESI Order, the 2023 logs ceased to be available long before that date.

24   [6] Plaintiffs' complaint about AWS instances is also groundless; the files downloaded to them were
     retained in other storage and produced to Plaintiffs. BG II Ex. 58 (2/10/25 Frederiksen Report)
25   ¶ 82 nn. 133, 135, 137.

26   [7] If Plaintiffs' argument is, instead, that Meta's copying of their works was not fair because Meta
     uploaded the unidentified works of *other* authors, that would be even further afield. Plaintiffs have
     no protectable interest in third party works, could not have been harmed by their alleged
27   distribution, and would not have standing to assert such a claim. By the same token, the repeated
     assertion in the IASTM Brief that Meta distributed *their* works or those of other third parties is
28   unsupported by evidence and irrelevant to the issues before the Court. Only Plaintiffs' claims are
     before the Court, and they have not shown distribution of *their* works or even part of them.

A.   Meta's Motion for Summary Judgment of Fair Use is Properly Directed to All of Its Copying of Plaintiffs' Works to Train and Develop Llama

As alleged in the TAC, Plaintiffs' claim for violation of § 106(1) is that, "*[t]o train Llama,* Meta downloaded and made several copies" of various datasets, including Books3, LibGen, Z-Lib, and Sci-Hub, which included the "Infringed Works," and that "Meta made copies of the Infringed Works *during the training process to develop Llama* without Plaintiffs' permission." TAC (Dkt. 407) ¶¶ 98, 100. Further, the TAC defines the "Infringed Works" as "the copyrighted books that Meta copied and used without permission *to train* Llama, regardless of where or how Meta downloaded or otherwise accessed the books." *Id.* ¶ 46. In other words, the only copying the TAC alleges constituted infringement under § 106(1) was Meta's copying to develop and train Llama. Accordingly, Meta moved for summary judgment that its "copying of Plaintiffs' works to *develop and train* large language models is fair use and does not infringe Plaintiffs' rights under 17 U.S.C. § 106(1)." Meta Mot. at 1 (Notice of Mot.); *see id.* at 2, 36, 37 (same). Meta's Motion does not "ignore[]" "other infringements" (Opp. at 10–11); no other "infringements" are alleged or at issue.

By the same token, all of the copying detailed at pages 12–14 of the Opposition was, even by Plaintiffs' telling, undertaken to develop and train Llama. Plaintiffs allege that an employee in France downloaded books from LibGen in 2022 to "determine 'if there [wa]s value' in copyrighted books as training data" for Llama, which was plainly part of the development process even if these copies ultimately were not used to train a Llama model released to the public. Opp. at 12. So, too, was use of works for "LLM research and evaluation, including 'assess[ing] the general knowledge and expressive abilities of models, and as a means of testing LLM's memorization." *Id.* at 12 n.9. It is standard practice (and essential to AI development) to perform experiments and analyses to assess the potential impact training data may have on the model. These analyses help AI researchers identify the types of data likely to improve model performance, as well those that may diminish it. Pineau Decl. ¶¶ 5–6; Kambadur Decl. ¶ 6; BG II Ex. 63 (Kambadur Dep.) 32:10–33:19, 35:17– 36:15, 117:13–118:14. Consistent with this research and development purpose, LibGen data downloaded in 2022 was used to determine whether and how use of the data would impact LLM performance, as measured using industry benchmarks. *Id.* And Plaintiffs do not explain how "cross-

referencing" content downloaded from LibGen in 2023 and Anna's Archive in 2024, which *was* used to train Llama, to assess whether it contained third party content (Opp. at 13–14), could possibly be an infringement of *Plaintiffs'* rights or require a separate fair use analysis. There is no allegation or evidence that such evaluations entailed making further copies of Plaintiffs' books.

In sum, all copying described in the TAC or Plaintiffs' Opposition was inarguably performed for purposes of developing and training Meta's transformative Llama models, and constitutes fair use as a matter of law. There are no instances of copying alleged by Plaintiffs that are not squarely addressed by Meta's Motion for summary judgment and subject to its fair use defense.[8]

B.   Bad Faith, If Relevant at All, Does Not Preclude Fair Use

Plaintiffs continue to focus heavily on Meta's downloading of works from online repositories, asserting that such conduct constitutes "bad faith." Opp. at 21–22. But the Supreme Court has not factored bad faith into its fair use analysis in any case since *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985). A decade later, the Court expressed uncertainty as to "the weight one might place on the alleged infringer's state of mind" and concluded that "[i]f the use is otherwise fair, then no permission need be sought or granted." *Campbell*, 510 U.S. at 585 n.18. And the Court has recently confirmed not only that there is cause for "skepticism" that "bad faith" should play "*any* role in a fair use analysis," but that such skepticism is "justifiable." *Oracle*, 593 U.S. at 33 (citing *Campbell* and Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1126 (1990) ("Leval")). Since *Oracle*, district courts hold that the concept of bad faith has little, if any, relevance to the fair use analysis. *See Teradyne, Inc. v. Astronics Test Sys., Inc.*, 2023 WL 9284863, at *17 n.17 (C.D. Cal. Dec. 6, 2023) (summary judgment of fair use); *Santos v. Kimmel*, 745 F. Supp. 3d 153, 165 (S.D.N.Y. Aug. 19, 2024) (same), *appeal argued*, No. 24-2196 (2d Cir. Mar. 24, 2025); *Thomson Reuters Enter. Centre GmbH v. Ross Intel. Inc.*, 2025 WL 458520, at *8 (D. Del. Feb. 11, 2025) ("[e]ven if relevant, bad faith would not move the needle").

To be sure, in *Harper & Row* the Supreme Court said that "[i]n evaluating character and

---

[8] Amici assert that the Court should distinguish between Meta's "reproductions via BitTorrent versus its subsequent reproductions to train the LLM," because there may be different "market harms," including harms from the alleged simultaneous distribution of Plaintiffs' works. IASTM Br. at 11–12. As discussed at length above, Plaintiffs have no evidence that Meta distributed any of their works, and this claim is not properly before the Court. *Supra*, at 4–7. Thus, the only "harm" Plaintiffs can properly attempt to show is from Meta's copying.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

purpose" it could not "ignore" the defendant's "stated purpose of scooping" its competitor by publishing verbatim excerpts of a "purloined" copy of an unpublished manuscript. 471 U.S. at 562–63. But the Court announced no categorical rule that a fair use must begin with an "authorized copy." Instead, it viewed theft of the manuscript as inseparable from the nature and purpose of the infringing use. As the Court explained, publication of those excerpts "had not merely the incidental effect but the *intended purpose* of supplanting the copyright holder's commercially valuable right of first publication." *Id.* at 562. In that sense, the use was not transformative—both the rightsholder and the defendant sought to be the "first publish[er] of a noted figure's copyrighted expression." *Id.* at 561; *cf. NXIVM Corp. v. Ross Inst.*, 364 F.3d 471, 479 (2d. Cir. 2004) (distinguishing *Harper & Row* where use of unpublished work was "quite plainly critical and transformative").

This case is miles away from *Harper & Row*. None of the challenged works at issue here was unpublished, nor does Meta seek to publish them. Unlike the article in *Harper & Row*, which preemptively published verbatim a substantial portion of an unpublished work, it is undisputed that Llama cannot be used as a substitute for reading Plaintiffs' books. And critically, the method by which Meta acquired Plaintiffs' books has no bearing on the nature and purpose of its use. Had Meta instead struck a deal with libraries to scan or download their books (as in *Google Books*), the resulting data would have been used in an identical way, Plaintiffs likewise would not have received any remuneration for those copies, and Meta's use to develop and train Llama would still be a quintessentially transformative one.[9] Thus, unlike in *Harper & Row*, the manner of acquisition in this case is irrelevant to whether Meta's use of the works was transformative and fair. *See* Leval, at 1126 (fair use "focus[es] not on the morality of the secondary user, but on whether her creation claiming the benefits of the doctrine is of the type that should receive those benefits[,]" primarily focusing "on whether the secondary use is productive and transformative and whether it causes excessive injury to the market for the original"); *see also* IP Profs' Br. at 2–4.[10]

As Plaintiffs acknowledge (Opp. at 4), the fair use doctrine *always* contemplates an

---

[9] Likewise, to borrow Plaintiffs' example (Opp. at 6), had Meta stolen physical copies of Plaintiffs' books from a bookstore to train Llama, the *store owner* would have a cause of action for the theft, but that conduct would not give rise to a *copyright* claim by Plaintiffs.

[10] The IASTM Brief (at 11) miscites *Google Books*, 804 F.3d at 208, for the proposition that the

unauthorized use (copying, display, distribution, etc.) that would be infringing but for its satisfaction of the statutory fair use factors. For example, in *Oracle*, "Google copied portions of the Sun Java API precisely," 593 U.S. at 30, in part because it did not want to comply with all of the terms of Oracle's standard license for its API, *see id.* at 8. "But since virtually any unauthorized use of a copyrighted computer program . . . would do the same," the relevant inquiry for fair use was not whether the copying was unauthorized, but rather whether "its use was consistent with that creative 'progress' that is the basic constitutional objective of copyright itself." *Id.* at 30.

The categorical claim that fair use must begin with an "authorized" copy, *Atari Games Corp. v. Nintendo of Am. Inc.*, 975 F.2d 832, 843 (Fed. Cir. 1992), is thus nonsensical—the point of the statutory protection is that the copying is unauthorized but nevertheless permitted. That is why the Ninth Circuit has repeatedly found fair use even when the use began with an unauthorized copy. *See* Meta Mot. at 16 & n.8 (discussing *Perfect 10*, 508 F.3d at 1164 n.8, and *Kelly v. Arriba Soft. Corp.*, 336 F.3d 811, 815 (9th Cir. 2003) (granting summary judgment)). Indeed, in *Connectix*, the Ninth Circuit held that the defendant's reverse-engineering of Sony BIOS software to create a competing game system constituted fair use "as a matter of law," 203 F.3d at 608, even though Sony had initially used "bootleg copies" of the BIOS "found on the Internet." *Sony Comput. Ent., Inc. v. Connectix Corp.*, 48 F. Supp. 2d 1212, 1219 (N.D. Cal. 1999). Thus, even if *Atari* remains good law in the Federal Circuit after *Campbell* and *Oracle*,[11] its overbroad statement that fair use requires "an authorized copy of a literary work," 975 F.2d at 843, cannot be correct, has never been endorsed by the Supreme Court (not even in *Harper & Row*), and is not the law in the Ninth Circuit.

Under the correct view of the law, then, it is clear that as long as Meta's use was ultimately fair, Meta would have been able to make an initial copy of the works without the rightholder's permission. Throughout their brief, Plaintiffs repeatedly use the terms "stolen," "pirated," and "illegal" to describe a work that has been copied or distributed without permission, but that is just

court "took pains to note that Google's initial acquisition of the copyrighted works was not itself an infringing act." However, no such finding (or discussion) appears anywhere in the decision. The cited page of the decision, from its "Background" section, merely noted that Google obtained digital copies of 20 million books through bi-lateral agreements with libraries, which absent author consent would undoubtedly be an "infringing act" but for fair use.

[11] Meta was unable to find a single case post-dating *Oracle* that relies on *Atari* for the proposition that fair use requires an authorized copy or that bad faith otherwise precludes a fair use defense.

a more sensational way of saying "unauthorized" or "infringing." That is so because a making a digital copy of a work without the rightsholder's permission does not, on its own, implicate any law other than the Copyright Act. Plaintiffs also try to refocus on the fact that a third party made an initial copy without permission, so as to ascribe that wrongdoing to Meta and taint the analysis as to whether its own use was fair. But it is difficult to see why the manner in which the third party acquired its copy should matter for purposes of *Meta's* liability under copyright law. While the third party may be liable for the original infringement, the fact that Meta made an unauthorized copy of an unauthorized copy seems of little importance—Meta's copy was going to be unauthorized either way. There is no reason for a different outcome merely because a third party made the initial copy without permission, instead of the defendant. Such a rule would have no correlation with the incentives an author has to create useful, transformative works.

Plaintiffs are seemingly asking the Court to create a rule that one can never make a fair use of an infringing copy prepared by someone else. But that would contravene *Perfect 10*, *Kelly*, and *Connectix*—all of which found fair use even though the defendants made use of works that were available on the Internet only because a third party had infringed them first. Plaintiffs' proposed rule also finds no support in *Harper & Row* or *Atari*—in both cases the copy used was "illegal" not because it was "infringing," but because its acquisition violated laws other than copyright. *See Harper & Row*, 471 U.S. at 562 (physically "purloined" manuscript); *Atari*, 975 F.2d at 836, 841–42 (fraud on Copyright Office). Taken to its logical extreme, Plaintiffs' proposed rule would mean that no one could write a fair use criticism or news article about an infringing work, both of which plainly serve a transformative purpose and would not substitute for the original. That is not the law.

Notably, although Plaintiffs describe their argument that "fair use cannot apply if the copyrighted work was illegally obtained in the first place" as "uncontroversial" (Opp. at 4), *even their own amici controvert it*. Their amici accept "the unremarkable proposition" that bad faith "is *not dispositive* to the fair use determination," and instead recognize that the relevant authorities support, at most, only the narrower proposition that bad faith acquisition should be "considered" under the first factor. *See* IASTM Br. at 12, 14. Here, as discussed above, the source of the copies Meta used to develop and train Llama has no bearing on whether Meta's use of those copies was

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1
2

transformative, and Plaintiffs and their amici do not explain how it could. Thus, even if considered, Meta's good or bad faith does not alter the balance of the fair use factors.

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

The many filesharing cases cited by Plaintiffs and amici are not to the contrary, because they involve the non-transformative distribution of music and video files for normal consumption, and thus could not satisfy the fair use factors irrespective of the provenance of the initial copy. *See* Meta Mot. at 33–36. Nor is there any merit to (or evidence for) Plaintiffs' speculation that a fair use finding in this case will "incentivize far-reaching piracy" and "incentivize LLM companies to support and defend shadow libraries." Opp. at 35. Whether use of materials on a filesharing site to train an LLM (or for any other purpose) is infringing turns on the particular facts of that use. The Court may only consider the specific use by Meta before it, *Goldsmith*, 598 U.S. at 533, which satisfies the fair use factors and is not infringing. If Plaintiffs are making a policy argument that the fair use doctrine should be constrained, and the scope of Plaintiffs' copyright monopoly expanded, to disallow use of "unauthorized" materials even for a fair use purpose, the Supreme Court has directed that such matters are for Congress to decide. *See Sony Corp. of Am. v. Univ. City Studios*, 464 U.S 417, 430–31 (1984) (describing the Court's "consistent deference to Congress when major technological innovations alter the market for copyrighted materials" and "reluctance to expand the protections afforded by the copyright without explicit legislative guidance").

18
19
20
21

The only relevant question before the Court is whether the balance of factors shows that *Meta* made transformative fair use of copies of Plaintiffs' works to develop and train Llama. Because the answer to that question is "yes," the fair use factors would weigh decisively in Meta's favor even if the Court were to find that the copies Meta used for that purpose were "unauthorized."

22

C.   The Fair Use Factors Weigh Decisively in Favor of Fair Use

23
24
25
26
27
28

On the core question presented, Meta's copying of datasets containing Plaintiffs' books to develop and train Llama furthers the objectives of copyright by enabling the creation of a new, transformative technology that serves a manifestly different purpose from Plaintiffs' books without any cognizable harm to Plaintiffs. Their cited evidence creates no triable issue as to fair use, and their arguments badly misconstrue the relevant authorities. The Court should rule that Meta's copying was fair use as a matter of law and grant Meta summary judgment on the § 106(1) claim.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1

2

1.   The purpose and character of Meta's use strongly favor fair use (Factor One)

3   Plaintiffs offer no rebuttal to Meta's showing that, under well-settled case law, Llama is

4   transformative because it does not "merely 'supersede[] the objects' of" Plaintiffs' works, but

5   plainly "adds something new, with a further purpose or different character." *Goldsmith*, 598 U.S.

6   at 527–29 (quoting *Campbell*, 510 U.S. at 579). It is undisputed that the purpose of Plaintiffs'

7   books is to be read, and they cannot be read using Llama. Meta Mot. at 15. The same was true in

8   *Google Books*. Google copied 20 million books to create its search tool, but the court had "no

9   difficulty" concluding at summary judgment that this was transformative, even though up to 16%

10 of each book could be read using the tool, because its purpose was to "make available significant

11 information *about those books*," not to reproduce them. *Authors Guild v. Google, Inc.*, 804 F.3d 202,

12 216–17, 223–24 (2d Cir. 2015) (emphasis in original).

13   Llama is even more transformative than Google's search tool. Its purpose is not to return

14 snippets of text from specific books, but to generate entirely new text based on statistical analysis

15 across its vast and varied training data. Meta Mot. at 10. Further, Plaintiffs' own evidence

16 establishes that, even subjecting Llama to intense adversarial prompting, the most expression it

17 could be made to output from Plaintiffs' books was, on average, a single passage (from the entire

18 book) of "roughly 50 tokens," or up to 50 words, Opp. at 21, far less than was found fair use in

19 *Google Books*. Plaintiffs' expert disclaimed that Llama could reproduce any significant percentage

20 of Plaintiffs' books. BG Ex. 24 (Lopes 3/7 Dep.) 237:16–19 ("Q: You are not offering any opinion

21 that Llama is able to reproduce, you know, any significant percentage of these books, correct? A:

22 Correct."); *see also* BG Ex. 23 (Lopes 2/13 Dep.) at 141:22–143:5 & BG II Ex. 56 (Lopes 3/7 Dep.)

23 at 223:16–224:11, 234:5–235:21, and 238:13–21 (only a few passages from any book outputted).

24   Plaintiffs half-heartedly refer to Meta's concerns about the *potential* for regurgitation, as

25 well as documents reflecting Meta's efforts to ensure that "models do not memorize and regurgitate

26 training data." Opp. at 20–21. These materials show that Meta sought to *avoid* infringement, not

27 that any meaningful regurgitation occurs. *See, e.g.*, Pritt Ex. 120 (Esiobu Dep.) 83:1–4 (testifying

28 that the rate of regurgitation was "extremely low … very close to zero"); Pritt Ex. 125 (LeCun Dep.)

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

129:11–131:22 (explaining that LLMs generally cannot "regurgitate the whole novel or even a chapter or even a paragraph probably," but may be able to output a "famous quote" if it "appear[s] thousands of times in a training set"). That intent to avoid infringement is why Meta deduplicated the training data before training Llama. Meta Mot. at 9, 38–39.

Plaintiffs posit repeatedly (without citing any evidence) that Meta "intentionally designed its LLMs to emulate the writing style of specific authors," and that this somehow creates a "factual dispute." Opp. at 16; *see also id.* at 17 (asserting that "when a Llama user asks the model to write a work of fiction in the style of Richard Kadrey, Llama will do it"); *id.* ("Meta produced fine-tuning data … intended to replicate the writing style of named Plaintiff Junot Diaz") (citing Pritt Ex. 112).[12] None of this helps Plaintiffs, because style is not protectable by copyright.

It is black-letter copyright law that "artistic styles" constitute ideas or concepts that are excluded from copyright protection under the idea-expression dichotomy. *See* 17 U.S.C. § 102(b); Meta Mot. at 12; *Mattel, Inc. v. MGA Ent., Inc.*, 616 F.3d 904, 913–17 (9th Cir. 2010) (copyright law precludes any "monopoly" over stylistic ideas and concepts); 2 Patry on Copyright § 4:14 (Mar. 2025 update) (style is "not protected"). This prohibition on protection for style specifically extends to "writing style." *Douglas v. Osteen*, 317 F. App'x 97, 99 (3d Cir. 2009); *see also Whitehead v. CBS/Viacom, Inc.*, 315 F. Supp. 2d 1, 11 (D.D.C. 2004) ("[S]tyle alone cannot support a copyright claim."); *Nesbitt v. Shultz*, 2001 WL 34131675, at *7 & n.4 (M.D. Pa. May 10, 2001) ("style[s] of writing" are unprotectable "ideas or concepts"). Plaintiffs cite no authority for their sideshow "style" argument, and it certainly does not raise a triable issue as to Llama's transformativeness.

Plaintiffs' final argument to sidestep the unavoidable conclusion that factor one weighs for Meta is that Llama does something *so different* from Plaintiffs' books that it has no "critical bearing" on them and, thus, is not a "traditionally transformative use[]." Opp. at 18–19 (agreeing that Llama

---

[12] Pritt Ex. 112 is a list of questions users might ask Llama, prepared to finetune the model on how to respond. For instance, in response to Plaintiff Klam's prompt to output text from his book, Llama said it would not do so for copyright reasons. BG Ex. 4. Teaching the model to respond that way required post-training. BG Ex. 47 (Scialom Dep.) 160:16–161:20, Ex. 48 (Touvron Dep.) 256:13–258:7. That Meta anticipated style-type questions does not show that Meta "intended" its models to replicate any Plaintiff's style, and none of Plaintiffs' cited exhibits shows that Llama can or would do so. In any event, Plaintiffs do not identify any Llama outputs that purportedly replicate their copyrighted expression, BG Exs. 1–3, 5, which is all that is protected.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   can "serve as a personal tutor," "assist with creative ideation, and help users to generate business

2   reports" among many other uses unrelated to Plaintiffs' books). This argument backfires.

3   The Supreme Court has repeatedly held that the examples of fair use enumerated in § 107

4   (*e.g.*, criticism, comment, scholarship, research) are "illustrative [] not limitative." *Campbell*, 510

5   U.S. at 577. Section 107 "set[s] forth *general* principles, the application of which requires judicial

6   balancing, depending upon relevant circumstances, including 'significant changes in technology.'"

7   *Oracle*, 593 U.S. at 19. The "more transformative the new work, the less will be the significance

8   of other factors …." *Campbell*, 510 U.S. at 579; *Goldsmith*, 598 U.S. at 531 ("a use that has a

9   distinct purpose is justified because it furthers the goal of copyright … without diminishing the

10   incentive to create"); *Google Books*, 804 F.3d at 214 ("The more the appropriator [uses] copied

11   material for new, transformative purposes, the more it serves copyright's goal of enriching public

12   knowledge and the less likely [] the appropriation will serve as a substitute for the original[.]").

13   Thus, Plaintiffs' suggestion that Llama's purpose and function are *too different* to qualify for fair

14   use protection is not only self-defeating, it turns copyright law on its head.

15   Plaintiffs also take language from *Goldsmith* out of context to argue that a use must have a

16   "critical bearing" on the underlying works to be transformative. Opp. at 19. That, too, is wrong.

17   The selected quotations are drawn from the Court's lengthy discussion of parody and other forms

18   of commentary, in which it noted that "parody needs to mimic an original to make its point,"

19   *Goldsmith*, 598 U.S. at 530, and, "[m]ore generally, when *commentary* has no critical bearing on

20   the substance or style of the original composition, … the claim to fairness in borrowing from [it]

21   diminishes accordingly[.]" *Id.* at 546. But Llama is not a parody; nor does its transformativeness

22   arise from any attempt to comment on Plaintiffs' works. Instead, it is transformative because it

23   serves a different purpose from Plaintiffs' books, which they now concede. That weighs heavily

24   in favor of fair use. *See, e.g.*, *Oracle*, 593 U.S. at 30–32 ("Google's use of the Sun Java API seeks

25   to create new products" was "transformative"); *Perfect 10*, 508 F.3d at 1165 (Google's display of

26   thumbnail images was transformative because its tool "provide[d] an entirely new use"); Meta Mot.

27   at 15–17 (citing additional cases).

28   Llama's "commercial purpose" is of little moment. As noted in Meta's Motion, "[m]any of

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   the most universally accepted forms of fair use, such as news reporting and commentary, … as well

2   as parody, are all normally done commercially for profit."   *Google Books*, 804 F.3d at 219.

3   Plaintiffs offer no explanation for why Llama's availability for both research and commercial

4   purposes might alter the fair use analysis here; it does not.   *See* Meta Mot. at 18 (citing numerous

5   cases finding commercial uses fair as a matter of law).

6   **2.   The "nature of the copyrighted work" favors fair use (Factor Two)**

7   Plaintiffs are wrong to suggest (Opp. at 22–23) that analysis under the second factor begins

8   and ends with the fact that Plaintiffs' books are creative, literary works. As *Google Books* observed,

9   the fact that certain of the plaintiffs' books were "factual" was not a "boost to Google's claim of fair

10   use," and conversely, "[i]f one (or all) of the plaintiff works were fiction," that would not have

11   "change[d] in any way [its] appraisal" that Google's coping of the books was fair. 804 F.3d at 220.

12   It is true, of course, that copyright protects only an "author's manner of expressing [] facts

13   and ideas." *Google Books*, 804 F.3d at 220. But Plaintiffs' argument (echoed by several amici) that

14   factor two weighs against fair use because Meta's initial copying of Plaintiffs' works captured not

15   only unprotected statistical information, but also their "creative expression," misses the mark. Opp.

16   at 22–23. In *Google Books*, Google copied the expression in millions of copyrighted books (and even

17   made up to 16% available for viewing in snippets). 804 F.3d at 216–17, 223–24. Likewise, in *Kelly*

18   and *Perfect 10*, all of the "creative expression" in the copyrighted images at issue was captured in the

19   initial indexing of websites, and that expression was viewable to users of the defendants' search tools,

20   albeit in "thumbnail," rather than full-size, form. *Kelly*, 336 F.3d at 815; *Perfect 10*, 508 F.3d at 1165.

21   And, of course, in any case involving criticism, commentary, or parody, some creative expression from

22   the underlying work must be evident in the secondary work. Yet, such uses are fair because they serve

23   a different purpose from, and do not substitute for, the original works. *See* Meta Mot. at 15–17.

24   Here, as in *Google Books*, Meta copies works to extract unprotectable information about the

25   words they contain. *See* Meta Mot. at 10, 22. This necessarily involves capturing the "expression"

26   in that work in the first instance, including information about the "arrangement of the words." Opp.

27   at 23. But Meta does not make these copies for purposes of enabling Llama to reproduce Plaintiffs'

28   expression, and that expression is not preserved in its code or model weights. Instead, the pre-training

**META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC**

process extracts statistical relationships between words and concepts across all of the vast training data; none of the original expression from any individual work remains in the model. *See* Ungar Decl. ¶¶ 11–14, 31, 37–38, 44; Meta Mot. at 9–10. Nor does Llama output that expression when prompted. No one can use Llama to read Sarah Silverman's description of her childhood, or Junot Diaz's story of a Dominican boy growing up in New Jersey, and Plaintiffs do not claim otherwise. Thus, Llama is an even clearer example of using a copy of a work to extract unprotectable information than the search tool in *Google Books*. Its statistical modeling of the relationships between words across millions of texts reflects no creative expression, much less Plaintiffs' protected expression.[13]

Plaintiffs are also wrong to dismiss the Ninth Circuit's analysis in *Connectix* as irrelevant. Although that case addressed the copying of computer code rather than books, it nevertheless drew a clear line between copying for the purpose of reproducing protected elements versus copying for the purpose of accessing unprotected functional elements. *Connectix*, 203 F.3d at 603–04. The second factor weighed in favor of fair use because Connectix's repeated copying of Sony's software was "necessary to gain access to the unprotected functional elements within [Sony's] program." *Id*. at 603. Moreover, in adjudicating whether the repeated copying was "necessary," the court declined to scrutinize the everyday decisions of the engineers, or to second-guess whether they could have made fewer copies during the engineering process. *Id.* at 605. So too here, what matters is that unprotected information about the words used in Plaintiffs' works could not be obtained or processed without first copying those works. *See* Meta Mot. at 22. "The copyright resulting from the Plaintiffs' authorship of their works does not include an exclusive right to furnish the kind of information about the works that [search tools or generative AI models] provide to the public." *Google Books*, 804 F.3d at 225.

Nor is it true that fair use for intermediate copying is limited to the software context. Opp. at 19. Although the need to make intermediate copies to accomplish a transformative use frequently arises in software cases, there is no distinction between software cases and cases involving other types of works. Meta's use may not, as in *Connectix*, begin with a piece of software, but its need for multiple copies is identical: to extract the unprotected information needed for pre-training

---

[13] Multiple amici assert that Llama, as well as models developed by other companies, are designed to mimic and reproduce the expression in their training material. Pls.' Profs.' Br. at 10–11; AAP Br.at 7–9; *see* CA Br. at 9–10. This speculation is not unsupported and contradicted by the record, and amici's characterizations about what third party models may do is not relevant or probative here.

1   Llama, Meta first needed to be able to access it in the datasets, which required a digital copy.[14]

2   Where, as here, a copy is made and used to create a transformative technology and defendant "does

3   not reveal that digital copy to the public[,]" *Google Books*, 804 F.3d at 221, "both the making of

4   the digital copies and the use of those copies to offer the [] tool [a]re fair uses." *Id*. at 217 (citing

5   *Author's Guild, Inc. v. HathiTrust*, 755 F.3d 87, 105 (2d Cir. 2014)).

6   It also matters greatly that Plaintiffs' works were previously published. "Published works are

7   more likely to qualify as fair use because the first appearance of the artist's expression has already

8   occurred." *Kelly*, 336 F.3d at 820; *see also Harper & Row*, 471 U.S. at 564. Even when a work is

9   creative in nature and reproduced in whole, its prior publication can mean that the second factor

10 "weighs only slightly in favor" of the rightsholder. *Kelly*, 336 F.3d at 820. Under the same rubric,

11 the combination of Meta's use to extract only non-protected aspects of Plaintiffs' works and the fact

12 that they were previously published mean that this factor weighs, even if slightly, in favor of fair use.

13   **3.   Meta's copying of entire books was reasonable (Factor Three)**

14   Plaintiffs fail to address any of the cases cited in Meta's Motion holding that copying an

15 entire work where "reasonable" or "necessary" to achieve the purpose of the fair use. *See Kelly*,

16 336 F.3d at 821 ("reasonable" to "copy each of Kelly's images as a whole" as copying less would

17 "reduc[e] the usefulness of the visual search engine"); *Sony Corp. of Am. v. Universal City Studios,*

18 *Inc.*, 464 U.S. 417, 449–50 (1984); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605,

19 613 (2d Cir. 2006) (granting summary judgment despite use of entire works); *A.V. ex rel.*

20 *Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 639–40, 642 (4th Cir. 2009) (same).

21   Plaintiffs do not claim that Meta took more than was necessary from their books to

22 accomplish its transformative purpose of developing an LLM capable of understanding how

23 language is used across many forms of text. That would fly in the face of the foregoing authorities.

24 Google could have copied fewer books or only portions of books, the defendant in iParadigms could

25 have extracted only parts of papers, and the defendant in *Bill Graham* could have used sub-portions

26 of the at-issue posters. In each instance, using less than the full work would have "reduc[ed] the

27 ───────────────
    [14] Many transformative uses do not require an intermediate copy. For instance, anyone can make
    a protected parody of a movie or song upon seeing or hearing it, without first downloading it to a

28 computer. But software products like Llama, or the game system in *Connectix*, can be developed
    only through use of intermediate digital copies.

**META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC**

usefulness" of the secondary work, *Kelly*, 336 F.3d at 821, while use of the whole work was reasonable for what defendant sought to accomplish. The same is true here.

Plaintiffs also give short shrift to the pronouncement in *Google Books* that what matters is "not so much 'the amount and substantiality of the portion used' in making a copy, but rather the amount and substantiality of what is thereby made accessible to a public for which it may serve as a competing substitute." 804 F.3d at 222 (emphasis omitted). The court thus found that Google "satisfie[d] the third factor test" because, "[w]hile Google *makes* an unauthorized digital copy of the entire book, it does not reveal that digital copy to the public." *Id.* at 221–22. Because the same is undisputedly true of Llama, Meta also satisfies the third factor test.

4.   Llama does not harm any cognizable market for Plaintiffs' works (Factor Four)

As Meta's Motion established (at 24–33), factor four weighs in favor of fair use because Llama does not substitute for Plaintiffs' works and causes them no cognizable harm, but does provide tremendous public benefits. Plaintiffs cannot overcome that showing.

a.   Any so-called "market" for LLM training data is not cognizable

Plaintiffs devote much of their discussion of factor four to Meta's efforts to acquire training data through agreements with third party publishers, while wholly ignoring the core issue: whether Plaintiffs are legally entitled to demand compensation for non-substitutive transformative use of their works. *Bill Graham Archives*, 448 F.3d at 614. They are not, rendering the entire discussion inapposite. *See* Meta Mot. at 26–27.

A copyright holder "cannot prevent others from entering fair use markets merely by developing or licensing a market for … transformative uses of its own creative work." *Tresóna Multimedia, LLC v. Burbank High Sch. Vocal Music Ass'n*, 953 F.3d 638, 652 (9th Cir. 2020) (quoting *Bill Graham Archives*, 448 F.3d at 614–15). Thus, in assessing the fourth factor, courts must determine whether the secondary use serves a "different market function" from the original. *McGucken v. Pub Ocean Ltd.*, 42 F.4th 1149, 1164 (9th Cir. 2022) (explaining that market harm and transformativeness are "linked"). Only harm occasioned by market substitution is cognizable (*id.*; *see Campbell*, 510 U.S. at 591); and market substitution does not encompass licensing for

transformative uses, even where they compete. *See Connectix*, 203 F.3d at 607–08; *see also HathiTrust*, 755 F.3d at 99 ("The only market harms that count are the ones that are caused because the secondary use serves as a substitute for the original, not when the secondary use is transformative[.]"). Were it otherwise, rightsholders could always overcome a fair use defense by pointing out the hypothetical possibility of licensing their work for the transformative use.

Plaintiffs acknowledge that "[t]he fourth factor is concerned with the threat of market substitution" (Opp. at 25), but seek to evade this limitation by arguing—without any reasoned basis—that Meta's copying is not transformative. *See, e.g.*, *id.* at 28 (describing Meta's copying of Plaintiffs' books to develop and train Llama as "a textbook example of" a "non-transformative" market substitution), 33 (copying was "[s]traightforward substitution"). Indeed, every decision cited by Plaintiffs concerns a *non*-transformative use bearing no relation to Meta's use here.

In *McGucken*, for example, the court rejected the fair use defense because the defendant's "use was in no way transformative—the [infringing] article used McGucken's photos to depict the ephemeral lake, which was exactly the purpose for which they were taken and exactly the function for which the photos had been licensed to other websites." 42 F.4th at 1153, 1161. The Second Circuit made the same finding in *Goldsmith*. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 11 F.4th 26, 50 (2d Cir. 2021) (use of photograph in magazine article about Prince was not transformative and had been licensed in the past for that exact use), *aff'd*, 598 U.S. 508.[15]

Plaintiffs' authorities are, thus, inapposite. As discussed above and in Meta's Motion, Llama is a highly transformative technology. Llama can help someone plan their day, solve a math problem, or summarize a meeting; but *unlike* the accused works in Plaintiffs' cited cases, it does

---

[15] Plaintiffs' other cases involve similarly non-transformative uses that directly substituted for the original works. *See Monge v. Maya Mags., Inc.*, 688 F.3d 1164, 1175 (9th Cir. 2012) ("photos were not physically or creatively transformed"); *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001) (non-transformative reproduction/distribution of music files), *as amended* (Apr. 3, 2001), *aff'd sub nom.*, 284 F.3d 1091 (9th Cir. 2002); *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 184 (2d Cir. 2024) (non-transformative distribution of books for reading); *On Davis v. Gap, Inc.*, 246 F.3d 152, 175 (2d Cir. 2001) ("not transformative" use of wearable art), *as amended* (May 15, 2001); *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 920 (2d Cir. 1994) (non-transformative copying of scientific journals to be read), *as amended* (July 17, 1995); *Ross Intel.*, 2025 WL 458520, at *8 (non-transformative use of Westlaw headings to create a non-generative AI model designed to reproduce them and do exactly what Westlaw does).

1   not reproduce Plaintiffs' works or their protected expression or substitute for reading them. *Supra*

2   pp. 12–13; Meta Mot. at 9–10.[16] The only market Plaintiffs identify is the licensing of their books

3   for the transformative purpose of training an LLM, which is not a purpose for which they were

4   written. BG Ex. 14. Under Ninth Circuit authority, this is not a cognizable market as a matter of

5   law. *Connectix*, 203 F.3d at 607–08; *see also Bill Graham Archives*, 448 F.3d at 615.

6   **b.   There is no market to license Plaintiffs' books for LLM training
                         and none is likely to form**

7

8   It is undisputed that Plaintiffs have never licensed any of their books as LLM training data

9   (whether individually, through their publishers, or as part of a collective), or even offered to do so.

10   BG Exs. 10–12. Even if markets for transformative uses were cognizable, there is no evidence

11   supporting the existence of a market for licensing *Plaintiffs'* works as training data—not in 2022

12   when Meta began training Llama 1 and not today.

13   This is not surprising. As explained in Meta's Motion (at 28), LLMs rely on astronomically

14   large volumes of text, of which any individual book provides an infinitesimally small and negligible

15   contribution to model performance. Ungar Decl. (Dkt. 496) ¶¶ 42–44. Plaintiffs' books thus have

16   insignificant and indeterminable economic value as training data, which precludes the formation of

17   a market to license them. Meta Mot. at 28. Plaintiffs provide no direct rebuttal to these facts.

18   Plaintiffs nevertheless insist that Meta "concedes there is an established market for

19   purchasing Plaintiffs' copyrighted books," and that, at a minimum, Meta deprived Plaintiffs of a

20   sale of each book to Meta. Opp. at 32–33. Meta made no such concession. Plaintiffs overlook

21   that Meta also could have purchased used copies of their books or partnered with libraries to scan

22   them (as in *Google Books*), neither of which would have resulted in remuneration to Plaintiffs.[17]

23

24   [16] Plaintiffs continue to cite the report of their expert, Dr. Spulber, for the unsubstantiated assertion
     that Llama harms Plaintiffs by "allow[ing] the creation of works that could compete with Plaintiffs'
     works." Opp. at 33 n.22, 34. As noted in Meta's Motion, he admitted at deposition that he was

25   unaware of any instance in which a Meta LLM has been used to generate any book, much less a
     competing one or one that was read as a substitute for a book of Plaintiffs. BG Ex. 25 (Spulber

26   Dep.) 245:7–248:1, 263:18–266:10. Moreover, creating competitive works (as when one author
     borrows ideas from another) is legally distinct from creating a market substitute. At most, Plaintiffs

27   are asserting that Llama might do the former—not the latter.

     [17] The Amicus Brief of Copyright Law Professors makes this same flawed argument, asserting that

28   "all human authors have had to pay something … for most [of] the works from which they have

**META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC**

This is a red herring in any event, because Plaintiffs do not contend that it would make any difference to the factor four analysis if Meta bought a new book at retail instead of using a copy procured elsewhere. *See Author's Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d 445, 462 (S.D.N.Y.), *aff'd in part, vacated in part*, 755 F. 3d 87 (2d Cir. 2014) (rejecting Plaintiffs' exact argument). Their claim is that it was infringement either way to use "copies of the Infringed Works during the training process to develop Llama" (TAC ¶ 100), and they cannot show that there is any market for that use.

Plaintiffs do not meaningfully reckon with the phenomenon of market failure—the reasons a market for a particular use does not arise—or the facts cited by Meta as likely to cause failure in the relevant market here. *See* Meta Mot. at 30–32. They do not dispute that LLMs require vast quantities of text (BG Ex. 23 (Lopes 2/13 Dep.) 20:19–22); Ungar Decl. ¶¶ 42–44); that individual works possess negligible value for this purpose (Sinkinson Decl. ¶¶ 56–57); and that rights to license books like Plaintiffs' are retained by authors, necessitating hundreds of thousands, if not millions, of costly and time consuming individualized negotiations. *Id.* ¶¶ 58–61, 69–70; BG Exs. 17, 18. As Meta explained, these factors alone render impracticable any potential market for licensing trade books as training data for LLM development. Meta Mot. at 30–32; Sinkinson Decl. ¶¶ 55–70.

The issue of market failure extends well beyond trade books. "[A]ll of the factors that undermine the viability of a market that includes licensing books to train LLMs are exacerbated for other categories of text training data." Sinkinson Decl. ¶ 70. No one could conceivably license all of the potentially copyright protected data required to train an LLM, including innumerable trade books, textbooks, scientific papers, news articles, and code repositories, as well as every blog post, essay, press release, email, etc. swept up in massive web crawls that comprise the vast majority of text data used to train the models. *Id.*; *see* Nayak Decl. ¶ 7. Even just identifying which materials are copyright protected and by whom would be an impossible task. In a scenario such as this, in which "prohibitive[]" costs "prohibit the formation of a viable market in the first place," factor four

---

learned[.]" Pls.' Profs.' Br. at 7. The Supreme Court says otherwise. *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) ("every idea, theory, and fact in a copyrighted work becomes instantly available for public exploitation at the moment of publication"). Any human can read and learn from a book, whether they paid for it or not. And they can also copy the book's expression for a fair use purpose.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   strongly favors fair use. *HathiTrust*, 902 F. Supp. 2d at 463–64 (emphasis omitted).

2   Unable to dispute these facts, Plaintiffs direct the Court's attention elsewhere. They spend

3   multiple pages cataloging various Meta documents and communications describing the company's

4   early exploratory outreach to publishers (Opp. at 26–31) and two drafts of a more recent forward-

5   looking data acquisition strategy. Pritt Exs. 97, 138. None of these documents evidences a market

6   for using *Plaintiffs'* books as training data or that such a market is likely to develop.

7   As Meta explained in its Motion (at 28), it explored negotiations with publishers in 2022

8   and early 2023 but quickly encountered obstacles that prevented transactions from being

9   consummated. While Plaintiffs speculate that Meta was unwilling to pay for books (Opp. at 28),

10   the record shows that the conversations failed to advance because publishers lacked the necessary

11   rights to license works (BG Ex. 34 (Choudhury Dep.) at 22:22–23:17; BG Ex. 31 (Boesenberg

12   Dep.) at 55:1–58:5, 152:21–153:13; BG II Ex. 62 (Choudhury Dep.) at 20:22–22:7), had too few

13   titles to make any transaction useful (BG Ex. 30 (Bell Dep.) at 131:21–133:2, 136:6–20; BG Ex.

14   45 (Nho Dep.) at 148:8–149:2), and were demanding unrealistic sums (BG II Ex. 61 (Boesenberg

15   Dep.) at 83:1–22, 154:7–155:3)—all hallmarks of market failure. Sinkinson Decl. ¶¶ 55–71.

16   As to Meta's more recent strategy documents, they reflect efforts to revisit exploration of

17   deals as a "surgical lever" to obtain access to relatively "small amount[s]" of necessary data that

18   cannot be secured through other means. *See* Pritt Ex. 138 at -888. Far from showing that a licensing

19   market for *Plaintiffs' books* is likely to form, these documents memorialize the same problems that

20   hampered Meta's earlier outreaches. They detail, among other things, that "no amount of

21   reasonable licensing budget can … get the tens of trillions of high quality tokens" necessary to train

22   an advanced LLM (Pritt Ex. 97 at -194); and that, despite the narrow focus on discrete categories

23   of data, potential partners lack necessary rights, are offering lower volumes of data than anticipated,

24   and are seeking unreasonably high fees. Pritt Ex. 138 at -889. To wit, the document indicates that

25   one textbook publisher apparently demanded $45 million for a 3-year term for access to 25,000

26   books, which was "not even sufficient" for Meta to conduct a study to evaluate the data. *Id.*

27   Plaintiffs and their amici also refer to several third-party data agreements, but only one, a

28

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

one-off agreement between ███████ and ██████████, is in the record.[18] That agreement is explicitly *not* a copyright license. Pritt Ex. 139 at -760 (providing that ████████ may use any delivered titles "as a matter of contract and not as a license under any IP rights that may exist in or to the Company Content[.]"). As such, it exemplifies a point made by Dr. Sinkinson (and not rebutted by Plaintiffs): that developers may value and pay for access to curated and/or non-public training data as a matter of contract, even in situations where licensing of intellectual property rights is not required because the materials are not copyrighted or the use is a fair one. Sinkinson Decl. ¶ 63 (describing the example of an agreement for access to legal opinions in the public domain).

Moreover, the circumstances surrounding that agreement only serve to reinforce Professor Sinkinson's market failure analysis. Sinkinson II Decl. ¶¶ 11–12. Negotiations began in the summer of 2023 and did not conclude until September 2024 (BG II Ex. 64 (███████ Decl.) ¶¶ 3–5), resulting in a deal that could, *at most*, deliver a total of 5,000 titles at a cost of $5,000 each. Pritt Ex. 139 at -761. Yet because ████████ did not own the rights to the works, it had to first negotiate consent from individual authors before it could deliver their books. *Id.* ¶ 6. In the end, ██████ received only roughly 1,800 titles, and no more are expected. *Id.* ¶¶ 7–8. In other words, ██████ invested over 1.5 years to obtain, conservatively, 0.001818% of the text data needed to train Llama 3, at a cost of over $9,000,000. Obtaining the number of tokens needed to train Llama 3, for example, would require a years-or even decades-long process of executing tens of thousands of similar agreements—which is infeasible and impractical and undermines any likely market.

Contrary to Plaintiffs' characterization, these facts show more than "that *some* transaction costs would accompany licensing efforts." Opp. at 30. The cause of market failure is that conditions render transaction costs impracticably high relative to the value exchanged. Sinkinson Decl. ¶¶ 55, 61, 68. Those conditions are demonstrated by the immense effort ██████ had to

---

[18] Other data agreements referenced by Plaintiffs and their amici are not properly before the Court. The AAP Brief (at 11–13) purports to describe "AI Licensing Deals" between developers and its members, but "[a]mici are not parties and cannot introduce evidence." *High Sierra Hikers Ass'n v. Powell*, 150 F. Supp. 2d 1023, 1045 (N.D. Cal. 2001) (striking extra-record evidence), *aff'd in part, rev'd in part sub nom.*, *High Sierra Hikers Ass'n v. Blackwell*, 390 F.3d 630 (9th Cir. 2004). Their references to "extra-record facts" are improper and should be stricken. *Id.* Beyond that, AAP does not purport to have personal knowledge of any of these so-called licenses, describe their terms, attach the agreements or offer foundation for them, or explain how the purported agreements might demonstrate a market for the licensing of *Plaintiffs'* books.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1

2

3

4

5

6

invest to obtain access to a statistically irrelevant set of books. Whether publishers such as ▮▮▮▮▮▮▮▮ can, in theory, act as intermediaries to attempt to negotiate licenses on behalf of a small set of rightsholders (Opp. at 30), is irrelevant when the end result cannot even begin to satisfy LLM development needs. Plaintiffs also fail to account for the fact that books are just a small portion of overall training data needs, and any such intermediary model is inapplicable to the millions of other materials contained in Llama training corpuses. Sinkinson Decl. ¶¶ 69–70.

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

Plaintiffs also cite an analysis by Meta of public information, unaccompanied by any testimony, to suggest that Meta "extensively surveyed the training data licensing agreements," and determined that its competitors have executed several text licenses that evidence a market for Plaintiffs' works. Opp. at 29; Pritt Ex. 97. But that document includes no confirmed details regarding the terms of any agreements, including whether they are, in fact, copyright licenses as opposed to agreements for data access or provision of real-time information, or include rights to reproduce licensed content as outputs, and Plaintiffs have not put any such agreements into evidence. Accordingly, their unsubstantiated characterization of extra-record third-party agreements of which they have no personal knowledge is entitled to no weight.[19] And to the extent third-party agreements concern news publications or online user content, those categories of data are readily distinguishable from trade books. Sinkinson II Decl. ¶¶ 5–10. Large news publications own their content and can thus provide ready access to timely information and a high volume of factually rich text dating back decades, whereas securing the same volume of data from trade books might require thousands of independent negotiations. *Id.* The same is true for large online platforms like Reddit, which can provide a large volume of diverse conversational text. *Id.* ¶ 10.

22

23

24

25

26

27

Finally, Plaintiffs' speculative analogizing between the growth of iTunes following the demise of Napster and what could occur in the data licensing market if the Court finds in Plaintiffs' favor, is similarly unsupported and should be disregarded. Plaintiffs cite no documents or expert opinion for this conclusion and instead rely on a single journal article, which is not admissible evidence. *See Valadez v. Cnty. of L.A.*, 2024 WL 3153207, at *10 (C.D. Cal. June 24, 2024)

28

---

[19] The same is true of the recent press release Plaintiffs cite. *See* Opp. at 30 & n.18. The article provides no indication of whether or what licenses are contemplated, is outside the record, and cannot be considered for its truth even if judicially noticed.

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   (declining to judicially notice "study" published in law review article); *In re Tesla, Inc. Sec. Litig.*,

2   477 F. Supp. 3d 903, 920 (N.D. Cal. 2020) (same).

3   In sum, Plaintiffs' efforts to manufacture a factual dispute over the likelihood of a market

4   developing are ineffective. Indeed, the Supreme Court has repeatedly found transformative fair use

5   notwithstanding the existence of failed licensing discussions over the exact work at issue. *Oracle*,

6   593 U.S. at 8; *Campbell*, 510 U.S. at 585 n.18. Plaintiffs can make no such showing here, and

7   Meta's use is even more transformative, warranting a fair use determination in its favor.

8   **c.   Llama's benefit to the public greatly outweighs any hypothetical
                            harm to the market for Plaintiffs' books**

9   Fair use serves as a "context-based check that can help to keep a copyright monopoly within

10 its lawful bounds" by ensuring authors are afforded no "more economic power than is necessary to

11 achieve the incentive to create" and cannot "us[e] copyright to stifle innovation" by others. Meta

12 Mot. at 13 (quoting *Oracle*, 593 U.S. at 21–22). Here, despite their protestations about how Meta

13 has harmed them, no Plaintiff claims that Llama's existence will prevent them from selling their

14 books or otherwise disincentivize them from writing. BG Ex. 13. Nor do they claim that the public

15 can access their work through Llama instead of buying (or borrowing) one of their books, i.e., the

16 substitutive harm recognized as "copyright's bête noire." *Goldsmith*, 598 U.S. at 528.

17   Instead, Plaintiffs effectively seek to extend the limited monopoly copyright affords to

18 cover a transformative use they never anticipated making or benefitting from. This would stifle

19 LLM innovation, set up insurmountable barriers to obtaining the vast volume and array of text

20 needed to train and develop useful models, and ultimately deprive the more than a billion people

21 who have downloaded Llama, and the countless others who have used it, of a groundbreaking new

22 technology that allows the creation of new, non-infringing content. *See* Meta Mot. at 11, 32

23 (detailing just some of the extraordinary public benefits and uses of Llama—none of which are

24 disputed by Plaintiffs). In other words, it would restrain, rather than promote, the progress of

25 science and the arts and the creation of new expression—the antithesis of copyright's objectives.

26 The Court should decline Plaintiffs' invitation to upend foundational copyright principles and hold

27 that Meta's use of their works to develop and train Llama was fair use as a matter of law.

28

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

IV.   THE COURT SHOULD GRANT SUMMARY JUDGMENT TO META ON THE DMCA CLAIM

Plaintiffs' claim that Meta removed CMI with intent to conceal infringement remains wholly unsupported. Faced with an unrebutted record showing that, consistent with common industry practice, Meta removed CMI and other repetitive text from training data to improve model performance (Meta Mot. at 37–40), Plaintiffs were required to proffer "evidence that would support a jury verdict." *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 256 (1986). Instead, they offer only unsupported inference and conjecture. They have no documents, testimony, or expert opinion showing that CMI was removed with knowledge or reason to know that it "will" conceal alleged infringement; indeed, they still lack a coherent theory of how CMI removal *could* do so. Meta Mot. at 39–40; *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675–76 (9th Cir. 2018). This claim cannot survive summary judgment. *Id.*; *Vaughn v. Teledyne, Inc.*, 628 F.2d 1214, 1220 (9th Cir. 1980).

First, despite nearly two years of litigation and multiple motions targeting dismissal of Plaintiffs' DMCA claims, Plaintiffs still do not offer a cogent explanation how CMI removal "conceals infringement" of their works. They cite no documents or testimony for an explanation. Nor did any of their experts opine on how the presence of CMI in a dataset would enable a user to query Llama to disclose the books upon which it was trained, while the absence of CMI would have prevented this. Mot. at 39–40. The most that Plaintiffs can muster is speculation that had CMI not been removed, "the © symbol[] could be regurgitated." Opp. at 37. But this makes no sense. The copyright symbol is not CMI, *see* 17 U.S.C. § 1202(c), and display of the symbol in Llama outputs would not reveal anything about whether Plaintiffs' works were used as training data. Their bare assertions, unsupported by evidence, cannot create a genuine issue precluding summary judgment for Meta. *Stevens*, 899 F.3d at 674–75 (affirming summary judgment because, among other things, plaintiff failed to aver how it would use CMI to "prevent or detect copyright infringement").

Moreover, there is nothing in the record to support that, *even if* CMI removal were somehow capable of concealing books present in the training data, Plaintiffs' ability to detect infringement could have been affected by it. As Meta has emphasized repeatedly, nothing was concealed from Plaintiffs. It is undisputed that Meta's public disclosure of its use of Books3 to train Llama 1 was the original basis of this lawsuit and, during discovery, Meta disclosed to Plaintiffs that it also used

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   Books3 to train Llama 2 and LibGen to train Llama 3. Meta Mot. at 39.

2   ***Second***, there is no evidence of scienter. Meta cited expert opinion, sworn testimony, and

3   a declaration from a member of the team that wrote the software for removing duplicative text from

4   LibGen, to establish that CMI removal was incidental to the commonplace effort of culling various

5   types of duplicative text from training data. Meta Mot. at 38–39. Mr. Bashlykov was explicit: "the

6   [data processing] team had no intent to hide or obscure any information" about Meta's training data

7   (Bashlykov Decl. ¶ 10), and he was unaware of any basis to believe that CMI removal "could have

8   the effect of concealing information about Meta's training data." *Id*. ¶ 11.

9   Again, Plaintiffs cite no contrary documents or testimony evidencing that training data

10   concealment is a possible, let alone known or desired, outcome of removing boilerplate copyright

11   information. Meanwhile, their own experts acknowledge that deduplication of training data helps

12   prevent overfitting (which can impair the ability of models to generalize) (BG Ex. 23 (Lopes 2/13 Dep.)

13   23:13–24), and that the software created by Meta to remove boilerplate information containing CMI

14   also removed repetitive text sequences that did not include CMI. BG Ex. 22 (Krein Report) ¶¶ 95–101.

15   Plaintiffs also do not dispute that removal of high-frequency text (i.e., text appearing thousands or

16   millions of times in a dataset) is common practice in processing training data. Ungar Decl. ¶¶ 65–69.

17   Lacking any affirmative evidence of scienter, and in the face of testimony establishing the

18   actual reasons why Meta filtered uninformative high-frequency text from datasets, Plaintiffs try to

19   manufacture a factual dispute with *ipse dixit* and strained inferences. For example, without any

20   supporting testimony, they contend that a jury could reasonably infer from Exhibit 56—a document

21   authored by Mr. Bashlykov (*see* BG II Ex. 60 (Bashlykov Dep.) 60:921) that describes the script

22   his team used—that because (i) the script also removed PII (personally identifiable information),

23   and (ii) PII removal was motivated by legal concerns, (iii) Meta's removal of CMI must have *also*

24   been motivated by (unstated) legal concerns. Opp. at 39, Pritt Ex. 56. But the document details no

25   such concerns, and the section Plaintiffs cite also describes "remov[al] [of] excessive new line

26   character[s]" and "lines that contain <8% [of] unique words" to address repetition more generally. *Id*.

27   Elsewhere, Plaintiffs suggest that concealment is evidenced by documents reflecting

28   discussions about finetuning to prevent Llama 2 from hallucinating about what it was trained on

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

1   (Opp. at 37, Pritt Ex. 149), and experiments conducted during development of Llama 3 to evaluate

2   whether adding attribution data to the pretraining corpus, including certain CMI, improves output

3   quality. Opp. at 40. But these documents do not concern CMI removal and, if anything, contradict

4   any inference of scienter—the former demonstrates the unreliability of the models to accurately

5   explain training data contents (Pritt Ex. 149), while the latter documents detail significant

6   uncertainty around the effect of adding "source metadata." *See* Pritt Exs. 137 at -970, 160.

7   Lastly, Plaintiffs turn to a newly-disclosed analysis of CMI removal in "B3G," the dataset

8   used by Meta that contains of portions of Books3 and Gutenberg, to argue that Meta's removal of

9   CMI from data in Books3, but not Gutenberg, implies a concealment motive. Opp. at 39. Plaintiffs'

10 B3G analysis constitutes undisclosed, untimely, and faulty expert testimony, improperly conducted

11 *by one of Plaintiffs' attorneys*, and should be stricken on those bases.[20] *See Jarrow Formulas, Inc.*

12 *v. Now Health Grp., Inc.*, 2012 WL 3186576, at *17 (C.D. Cal. Aug. 2, 2012) (excluding

13 undisclosed and untimely expert declaration despite claim that declaration was based on personal

14 knowledge), *aff'd*, 579 F. App'x 995 (Fed. Cir. 2014). But even if the Court were to entertain a

15 belated expert report prepared by a party attorney, it is not probative of concealment. Mr. Butterick

16 at best avers that a small fraction of works in the Gutenberg portion of B3G contain copyright

17 notices. Butterick Decl. ¶¶ 12–15. This is insufficient to create a plausible inference that Meta

18 intentionally removed CMI from Books3 to conceal infringement, not least because Meta *publicly*

19 *disclosed* its use of Books3 to train Llama 1. Meta could not have known or had reason to know

20 that CMI removal would conceal something that it conspicuously announced.

21

22 [20] Plaintiffs understood that analysis of CMI was the province of experts, as Dr. Lopes opined in
    her opening report that Meta removed CMI from copies of their books in the Books3 and LibGen
    datasets. Pritt Ex. 147. *None* opined about the presence of CMI in the Gutenberg portion of the
23 B3G dataset, even though it was produced in May 2024. Butterick Decl. (Dkt. 519) ¶ 9. Plaintiffs
    are now attempting to evade expert disclosure requirements "through the simple expedient of
24 proffering an expert in lay witness clothing." Fed. R. Evid. 701 advisory committee's notes to 2000
    amendments. Mr. Butterick's untimely expert opinion is also improper under Fed. R. Evid. 702,
25 as he does not establish that he is qualified to offer the opinions reflected in his declaration and
    does not disclose the methodology he used to search through tens of thousands of Gutenberg entries.
26 And even as they turn to Mr. Butterick for this improper analysis, they did *nothing* to show that the
    Llama models can identify books in Gutenberg (from which CMI was supposedly not removed) as
27 books on which they were trained, while books in Books3 that supposedly were subject to CMI
    removal cannot be so identified, rendering the analysis he does provide irrelevant and non-probative
28 as to concealment. In any case, it would be prejudicial to allow Plaintiffs to rely on Mr. Butterick's
    untimely declaration, to which Meta's experts had no opportunity to respond.

Third, Plaintiffs do not engage at all on the issue of their lack of statutory and Article III standing on this claim. *See* Meta Mot. at 40. They articulate no injury caused by Meta's alleged concealment effort because there is none. As noted, Meta publicly disclosed that it used Books3 to train Llama 1 and, during discovery, it disclosed to Plaintiffs that it also used Books3 for Llama 2 and portions of Libgen for subsequent models. Meta Mot. at 39. Moreover, only one of the Plaintiffs appears to have ever used Llama, and none testified that they had used Llama in an attempt to discern whether Meta trained on their works. BG II Ex. 54. Even if Plaintiffs could show that Meta removed CMI with culpable scienter and that such removal did, in fact, hide what books Llama was trained on, Plaintiffs *would still lack standing* because nothing was concealed from them. Such a scenario describes a statutory violation without a corresponding concrete injury, which is exactly what the Supreme Court deemed insufficient to support Article III standing in *TransUnion LLC v. Ramirez*. 594 U.S. 413, 426–27 (2021) (explaining that "an injury in law is not an injury in fact" and that "[o]nly those plaintiffs who have been *concretely harmed* by a defendant's statutory violation may sue"). Here too, Plaintiffs lack standing, warranting judgment for Meta.

Finally, we are aware of no case, and Plaintiffs cite none, where a 1202 claim was allowed as to a non-infringing fair use. Since Meta's use of Plaintiffs' works was fair, their DMCA claim fails for the additional reason that there was no infringement to conceal.

V.   CONCLUSION

Meta's copying of Plaintiffs' works to train Llama constitutes fair use under 17 U.S.C. § 107, and Plaintiffs have failed to come forward with evidence to create a genuine issue for trial on their claim that Meta violated 17 U.S.C. § 1202. Accordingly, Meta respectfully asks the Court to grant summary judgment for Meta on Plaintiffs' claims under 17 U.S.C. §§ 106(1) and 1202.
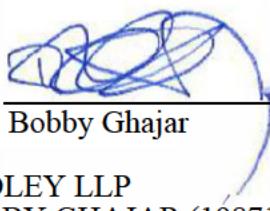
META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC

Dated: April 17, 2025

COOLEY LLP

By: _____
Bobby Ghajar

COOLEY LLP
BOBBY GHAJAR (198719)
(bghajar@cooley.com)
TERESA MICHAUD (296329)
(tmichaud@cooley.com)
COLETTE GHAZARIAN (322235)
(cghazarian@cooley.com)
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400

COOLEY LLP
MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
MATTHEW BRIGHAM (191428)
(mbrigham@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
JUAN PABLO GONZÁLEZ (334470)
(jgonzalez@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000

COOLEY LLP
PHILLIP MORTON (pro hac vice)
(pmorton@cooley.com)
COLE A. POPPELL (pro hac vice)
(cpoppell@cooley.com)
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004-2400
Telephone: (202) 842-7800

CLEARY GOTTLIEB STEEN &
HAMILTON LLP

By: _____
Angela Dunning

CLEARY GOTTLIEB STEEN &
HAMILTON LLP
ANGELA L. DUNNING (212047)
(adunning@cgsh.com)
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4131

PAUL, WEISS, RIFKIND, WHARTON &
GARRISON LLP
KANNON K. SHANMUGAM
(pro hac vice)
(kshanmugam@paulweiss.com)
WILLIAM T. MARKS (pro hac vice)
(wmarks@paulweiss.com)
2001 K Street, NW
Washington, DC 20006
Telephone: (202) 223-7300

PAUL, WEISS, RIFKIND, WHARTON &
GARRISON LLP
ANNA STAPLETON (337686)
535 Mission Street, 25th Floor
San Francisco, CA 94105
Telephone: (415) 432-5000

META'S REPLY ISO MOT. FOR PARTIAL
SUMMARY JUDGMENT
3:23-CV-03417-VC